

**DATABASE AND ANALYTICAL TOOL DEVELOPMENT  
FOR THE MANAGEMENT OF DATA  
DERIVED FROM US DOE (NETL) FUNDED  
FINE PARTICULATE (PM<sub>2.5</sub>) RESEARCH**

**SEMI-ANNUAL  
TECHNICAL PROGRESS REPORT**



**Submitted to:**

U. S. Department of Energy  
National Energy Technology Laboratory  
Pittsburgh, PA 15236

**DOE AWARD NUMBER: DE-FC26-02NT41476**

# **DATABASE AND ANALYTICAL TOOL DEVELOPMENT FOR THE MANAGEMENT OF DATA DERIVED FROM US DOE (NETL) FUNDED FINE PARTICULATE (PM<sub>2.5</sub>) RESEARCH**

## **Semi-Annual Technical Progress Report**

**REPORT PERIOD START DATE:** August 12, 2003

**REPORT PERIOD END DATE:** February 11, 2004

**PRINCIPAL AUTHORS:** Robinson P. Khosah, Ph.D.  
**Advanced Technology Systems, Inc.**  
Charles G. Crawford  
**Advanced Technology Systems, Inc.**

**DATE REPORT ISSUED:** February 2004

**DOE AWARD NUMBER:** DE-FC26-02NT41476

**CONTRACTOR:** **Advanced Technology Systems, Inc.**  
639 Alpha Drive – RIDC Park  
Pittsburgh, PA 15238-2819

**SUBCONTRACTORS:** **Ohio University**  
E317 Grover  
Ohio University  
Athens, OH 45701  
**Texas A&M University – Kingsville**  
MSC213  
Texas A&M University - Kingsville  
Kingsville, TX 78363

**ATS PROJECT NO:** 01-050-S&T-P

## ABSTRACT

Advanced Technology Systems, Inc. (*ATS*) was contracted by the U. S. Department of Energy's National Energy Technology Laboratory (DOE-NETL) to develop a state-of-the-art, scalable and robust web-accessible database application to manage the extensive data sets resulting from the DOE-NETL-sponsored ambient air monitoring programs in the upper Ohio River valley region. The data management system was designed to include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions. The project is being conducted in two phases. Phase One includes the following tasks: (1) data inventory/benchmarking, including the establishment of an external stakeholder group; (2) development of a data management system; (3) population of the database; (4) development of a web-based data retrieval system, and (5) establishment of an internal quality assurance/quality control system on data management. Phase Two, which is currently underway, involves the development of a platform for on-line data analysis. Phase Two includes the following tasks: (1) development of a sponsor and stakeholder/user website with extensive online analytical tools; (2) development of a public website; (3) incorporation of an extensive online help system into each website; and (4) incorporation of a graphical representation (mapping) system into each website. The project is now into its eighteenth month of development activities.

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

# TABLE OF CONTENTS

ABSTRACT .....	i
LIST OF GRAPHICAL MATERIALS .....	ii
EXECUTIVE SUMMARY .....	i
I. INTRODUCTION .....	1
A. Project Goals and Objectives .....	1
B. Project Phase Development .....	2
II. EXPERIMENTAL .....	3
A. Phase One Tasks .....	3
Task 1.1 – Data Benchmarking/Inventory for Database Applications .....	3
Task 1.2 - Develop Data Management System Architecture .....	3
Task 1.3 - Population of Database .....	4
Task 1.4 - Develop Web-Based Retrieval System.....	4
Task 1.5 - Develop QA/QC System .....	7
B. Phase Two Tasks .....	7
Task 2.1 - Develop Stakeholder Website .....	7
Task 2.2 - Develop Public Website.....	9
Task 2.3 - Develop Online Help Feature .....	9
Task 2.5 - Performance Testing .....	11
III. RESULTS AND DISCUSSION .....	11
IV. CONCLUSION.....	12
V. REFERENCES .....	13
VI. BIBLIOGRAPHY .....	14
VII. LIST OF TERMS, ACRONYMS AND ABBREVIATIONS .....	15

## LIST OF GRAPHICAL MATERIALS

1. Criteria Selection Controls.....	5
2. Query Results Display.....	5
3. XML Data .....	5
4. Query Tab Control.....	6
5. Output Type Selection Control .....	6
6. Download Help.....	10
7. Example Client-Side Toolbars.....	10
8. Example Box-Whisker Plot with Data Table .....	11

## EXECUTIVE SUMMARY

Advanced Technology Systems, Inc. (*ATS*) was contracted by the U. S. Department of Energy's National Energy Technology Laboratory (DOE-NETL) in August, 2002, to develop a state-of-the-art, scalable and robust web-accessible database application to manage the extensive data sets resulting from ambient air monitoring programs in the upper Ohio River valley region that have been sponsored by DOE-NETL.

Research projects sponsored by DOE-NETL collected large amounts of data on PM<sub>2.5</sub> and other air pollutants at ambient monitoring sites in the upper Ohio River valley region between 1999 and 2003. Extensive monitoring sites have been operated by DOE-NETL and its contractors in Pittsburgh, PA (two (2) urban sites), Holbrook, PA (rural site), South Park, PA (suburban site), and Steubenville, OH. Less-extensive monitoring sites have been operated in six other locations in PA, OH and WV. The main objectives of the current effort are to gather the data from all these monitoring sites into a common database, and to develop analytical tools that will make the data easily accessible to researchers and the public via the Internet.

In addition to the data collected by DOE-NETL and its contractors, the database will include, to the greatest extent possible, ambient air data collected by other agencies in the upper Ohio River valley region, such as the U.S. EPA, Pennsylvania Department of Environmental Protection (PA-DEP), West Virginia Division of Environmental Protection (WV-DEP), Ohio EPA, and the Allegheny County Health Department (ACHD). Although emphasis is being placed on the upper Ohio River valley region, the database may also include data collected at other DOE-NETL sponsored sites outside the region, such as sites operated by the Tennessee Valley Authority in the Great Smoky Mountains and by the Southern Research Institute in North Birmingham, AL. The database and analytical tool development effort is also being coordinated, to the extent possible, with a similar effort by U.S. EPA to develop a relational database for data collected at its "PM Supersites". This coordination will ensure that the database and analytical tools produced under the DOE-NETL effort will be readily accessible to a wide variety of stakeholders.

The data management system will include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

The project is being conducted in two phases. The entire project has been divided into ten primary tasks and those have been segmented into two primary phases. The project is now into its eighteenth month of development tasks and Phase Two began in August 2003. Phase One consisted of design and specification tasks related to designing, implementing and populating the primary database that will house the collected data. Phase Two consists of tasks involving the design, implementation and testing of both website interfaces along with any analytical tools and features integrated into the project's websites.

## **I. INTRODUCTION**

Advanced Technology Systems, Inc. (*ATS*) was contracted by the U. S. Department of Energy's National Energy Technology Laboratory (DOE-NETL) in August, 2002, to develop a state-of-the-art, scalable and robust web-accessible database application to manage the extensive data sets resulting from ambient air monitoring programs in the upper Ohio River valley region that have been sponsored by DOE-NETL.

Research projects sponsored by DOE-NETL collected large amounts of data on PM<sub>2.5</sub> and other air pollutants at ambient monitoring sites in the upper Ohio River valley region between 1999 and 2003. Extensive monitoring sites have been operated by DOE-NETL and its contractors in Pittsburgh, PA (two (2) urban sites), Holbrook, PA (rural site), South Park, PA (suburban site), and Steubenville, OH. Less-extensive monitoring sites have been operated in six other locations in PA, OH and WV. The main objectives of the current effort are to gather the data from all these monitoring sites into a common database, and to develop analytical tools that will make the data easily accessible to researchers and the public via the Internet.

In addition to the data collected by DOE-NETL and its contractors, the database will include, to the greatest extent possible, ambient air data collected by other agencies in the upper Ohio River valley region, such as the U.S. EPA, Pennsylvania Department of Environmental Protection (PA-DEP), West Virginia Division of Environmental Protection (WV-DEP), Ohio EPA, and the Allegheny County Health Department (ACHD). Although emphasis is being placed on the upper Ohio River valley region, the database may also include data collected at other DOE-NETL sponsored sites outside the region, such as sites operated by the Tennessee Valley Authority in the Great Smokey Mountains and by the Southern Research Institute in North Birmingham, AL. The database and analytical tool development effort is also being coordinated, to the extent possible, with a similar effort by U.S. EPA to develop a relational database for data collected at its "PM Supersites". This coordination will ensure that the database and analytical tools produced under the DOE-NETL effort will be readily accessible to a wide variety of stakeholders.

The data management system will include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

The project is being conducted in two phases. The project is now into its eighteenth month of development activities.

### **A. Project Goals and Objectives**

The main objective of this database development effort is to gather the data, acquired from all these monitoring sites, into a common database, and to develop analytical tools that will make the data easily accessible to researchers and the public via the Internet.

The proposed data management system will include a web-based user interface that will allow easy access to the data by the scientific community, policy- and decision-makers, and other interested stakeholders, while providing detailed information on sampling, analytical and quality control parameters. In addition, the system will provide graphical analytical tools for displaying, analyzing and interpreting the air quality data. The system will also provide multiple report generation capabilities and easy-to-understand visualization formats that can be utilized by the media and public outreach/educational institutions.

## ***B. Project Phase Development***

The project is being conducted in two phases. The first phase includes data inventory, benchmarking and database population tasks, as well as the development of data management architecture, a web-based retrieval system and an internal QA/QC system. A progress summary for Phase One is shown in the following table:

<b>Task #</b>	<b>Description</b>	<b>Planned completed %</b>	<b>Actual completed %</b>
1.1	Data Inventory/Benchmarking for Database Applications	100%	95%
1.2	Develop Data Management System Architecture	100%	95%
1.3	Population of Database	100%	65%
1.4	Develop Web-based Retrieval System	100%	90%
1.5	Develop Internal QA/QC System	N/A	N/A

N/A – on-going activity.

The second project development phase is currently underway. The tasks involved in this phase include the development of a stakeholder-specific website, a publicly accessible website and an online help feature. This phase also includes the development of special analysis tools to provide a graphical representation of the data and, of course, a series of performance tests designed to provide the best possible data management solution.

A progress summary for Phase Two is shown in the following table:

<b>Task #</b>	<b>Description</b>	<b>Planned completed %</b>	<b>Actual completed %</b>
2.1	Develop Stakeholder Website	100%	15%
2.2	Develop Public Website	100%	15%
2.3	Develop Online Help Feature	50%	5%
2.4	Provide Graphical Representation of Data	25%	5%
2.5	Performance Test	25%	0%



## II. EXPERIMENTAL

### A. Phase One Tasks

#### Task 1.1 – Data Benchmarking/Inventory for Database Applications

Any database application development effort requires some knowledge of the types and number of data contained in the resulting database. In addition to knowing this information, it is also wise to investigate or benchmark existing applications and development efforts that are similar in design or nature as the application being developed. Therefore, *ATS* proposed to conduct benchmarking investigations of existing projects, activities and applications prior to embarking on this project, as well as evaluate and quantify the data destined for usage with this application.

Several items were described in detail within the first and second Semi-Annual Technical Reports for this project. Those included the CARB Data Management Project in California and the EPA Supersite Database Development Project.

During Phase Two of this project, team members traveled to St. Louis, MO, to attend the Regional Planning Organization (RPO) Conference. The RPOs have been established by U. S. EPA to evaluate technical information to better understand how their national park and wilderness areas (Class I areas) across the country are impacted by PM<sub>2.5</sub> and related pollutants, and to pursue the development of regional strategies to reduce emissions of particulate matter and other pollutants leading to regional haze. It is anticipated that the RPO members will be among the most important users of the database and analytical tools developed under this project. Discussions were held with key participants at the conference to discuss potential needs for analysis tools and online database resources. Additionally, team members made inquiries regarding an existing PM data analysis software package titled Ambient Monitoring Data Analysis Software (AMDAS), and met with developers for the Visibility Information Exchange Web System (VIEWS) to discuss methods of obtaining, sharing and delivering data to end users.

The project team provided a presentation for the Data Analysis workshop attendees at the conference and received very favorable feedback. This presentation focused on the planned integration of analysis tools and methods of linking these tools to multiple datasets to broaden the potential user base.

The VIEWS project development team had a small computer lab available for researchers attending the conference to test-drive their web interface. Many of these researchers were questioned regarding their current toolsets in use for analyzing PM data. Specifically, questions were asked regarding AMDAS, to which several said they were aware of AMDAS, but have not devoted time to learn it because it is being discontinued due to compatibility issues with the parent application, S-PLUS, a statistical analysis application that allows AMDAS to operate as a *plug-in* module.

#### Task 1.2 - Develop Data Management System Architecture

The first semi-annual technical report contained detailed information regarding the design of the data management system architecture. To summarize, a system has been developed

using MS SQL Server 2000 Enterprise Edition, MS Windows 2000 Advanced Server and external hardware, to provide the data management system architecture for this project.

A series of database objects and scripts have been constructed to ease all software development tasks and to accommodate expansion of the system to accommodate more users and data.

The second semi-annual technical report contained detailed descriptions of *stored procedures*, or static queries that are stored within the database structure.

### **Task 1.3 - Population of Database**

A series of conversion scripts have been created to reformat the supplied data files into the application-specific format. Some data files have been submitted that have been formatted to comply with the published NARSTO standards. Since the NARSTO format utilizes specific headings and tables that do not directly correspond with the project-specific database structure, these files cannot be directly imported into the database. The conversion scripts allow the project team to convert this data into a suitable format that will allow the submitted data to be automatically merged with the existing data in the database.

The database structure, although complete in its present form, could undergo multiple changes and alterations while developing the database population source code. Once the population code has been completely tested, a final copy of all data will be processed into the database from the static, offline datasets stored on the server and submitted to the project team. The database population software is being coded to include a data tracking system so data quality will be more manageable.

The datasets that have been submitted to the project team for inclusion in this database have been used to establish a series of support datasets, including static information such as data collection protocols and parameter description codes which are required for NARSTO compatibility. This data has been prearranged in text files and Data Transformation Services (DTS) scripts have been created to create the database structure and then import this support data in a specified sequence.

### **Task 1.4 - Develop Web-Based Retrieval System**

Once the data is transferred to the database, users specify which data is to be retrieved through the Standard Query Builder Interface. This interface provides resulting datasets in a tabular format.

Until the database itself is made publicly available, static datasets have been provided via HTTP protocol. Users complete a criteria selection process to download these files. This process is similar to, but more limited than, the Standard Query Builder Interface. Users select the type of data desired from each collection site's parameter list controls, and then click on the "Download Files" button above the selection list controls, to retrieve their data files.

Download Files

Site	Select Parameters (Use Ctrl-click for selecting multiple)
Holbrook	Summary Mass Data
	SFS Mass Data
	SFS Speciation(Detailed)
	TEOM 2.5 Data
Lawrenceville	Summary Mass Data
	SFS Mass Data
	SFS Speciation(Detailed)
	TEOM 2.5 Data
Schenley Park	Surface TEOM GAS
	Continuous (SO4,NO3)
	Meteorological
	Denuded Organic Data
Morgantown	Summary Mass Data
	SFS Mass Data

### Criteria Selection Controls

These data files appear as hyperlinks with textual descriptions of their contents, so users know what they are downloading before they actually start downloading an individual data file.

Step 3) Use Prev and Next links to scroll through the Pages .  
 Step 4) Save the Excel Files To Your Computer By Right Clicking the Excel Image Button and "Save Target As" TO your computer.

<< Prev Next >> Page: 1/2 Pages

RIGHT CLICK ON EXCEL IMAGE LINK IN THE RIGHT COLUMN AND CLICK "SAVE TARGET AS" TO SAVE TO YOUR COMPUTER	
NETWORK NAME: Upper Ohio River Valley Project    SITE NAME: Holbrook, PA Parameters: Summary Mass Data Time Period: 01/01/2001 To 08/08/2001 File Size: 102.91 KB	 <a href="#">Right Click Here and Save Target As</a>
NETWORK NAME: Upper Ohio River Valley Project    SITE NAME: Holbrook, PA Parameters: Summary Mass Data Time Period: 01/22/2000 To 08/08/2001 File Size: 86.02 KB	 <a href="#">Right Click Here and Save Target As</a>
NETWORK NAME: Upper Ohio River Valley Project    SITE NAME: Holbrook, PA Parameters: Summary Mass Data Time Period: 02/17/1999 To 03/07/1999 File Size: 80.38 KB	 <a href="#">Right Click Here and Save Target As</a>
NETWORK NAME: Upper Ohio River Valley Project    SITE NAME: Holbrook, PA Parameters: Summary Mass Data Time Period: 02/17/1999 To 06/29/1999 File Size: 81.41 KB	 <a href="#">Right Click Here and Save Target As</a>
NETWORK NAME: Upper Ohio River Valley Project    SITE NAME: Holbrook, PA Parameters: Summary Mass Data Time Period: 03/13/1999 To 06/29/1999 File Size: 83.97 KB	 <a href="#">Right Click Here and Save Target As</a>
NETWORK NAME: Upper Ohio River Valley Project    SITE NAME: Holbrook, PA Parameters: Summary Mass Data Time Period: 04/30/2000 To 01/01/2001 File Size: 104.96 KB	 <a href="#">Right Click Here and Save Target As</a>

### Query Results Display

The textual descriptions are obtained by parsing the file names, which contain site, network, parameter and versioning information, which is associated with full-text descriptions within a statically defined XML data file.

```
<?xml version="1.0" encoding="utf-8" ?>
<filetypes>
<abbreviation name="UORVL" full="Lawrenceville, PA"/>
<abbreviation name="PAQS" full="Pittsburgh Air Quality Study"/>
<abbreviation name="UORVP" full="Upper Ohio River Valley Project"/>
<abbreviation name="UORVH" full="Holbrook, PA"/>
<abbreviation name="UORVM" full="Morgantown, PA"/>
```

### XML Data

## AD HOC QUERY SYSTEMS

The primary purpose of this project is to provide dynamic capabilities to this data retrieval system. Researchers need access to an ad hoc query system to build their own datasets, or to merge them with other datasets from multiple sources. The design of the

query interface is the most critical aspect to consider when designing a web-based data retrieval system. It should provide the user an efficient means of deriving the output required from the database, without needing to understand the inner structure of the database or requiring the technical knowledge for writing Structured Query Language (SQL) queries and stored procedures.

### ***Standard Query Builder Interface***

The Standard Query Builder Interface provides a logical flow in the data filter selection process so that the user can easily select subsets of data that they wish to view. Therefore, the user selects from a range of input criteria as well as defines the result sets based on user-specified output criteria. Input criteria will include selecting date ranges, collection site locations, monitoring networks, air quality parameters, collection principles, etc. Output criteria define what columns the user wishes to include in the resulting dataset, and include items such as date and time values, air quality parameters, durations, frequencies, and metafiles.

The Standard Query System provides an intuitive interface for accessing the database. The Query System works by building a SQL Query based on the choices the user makes during the selection process.

The sequential querying process is implemented by the Query Tab Control, displayed below. The Red bar under the Tabs displays the status of the querying process. The user can change the selection made at any stage by clicking on the particular tab.

#### **Query Tab Control**

1. Select Sites	2. Select Time-Period	3. Select Parameters	4. Select Output-Format	5. Finalize Query
Completed 3/5 Steps				

The Selection Process itself is split into 5 Stages displayed by the tabs.

1. Site Selection
2. Date Period Selection
3. Parameter Selection
4. Output Format Selection
5. Query Finalization.

Once these steps are completed, the users query is processed at the server and the resulting data is returned to the client's browser in a variety of formats. The project team will be working with stakeholders to refine the *Standard Query Builder Interface*.

#### **SELECT OUTPUT FORMAT**

Select Output Formats

- ☒ MS Excel Document File
- ☒ Comma Seperated Value File
- ☒ Tab delimited text file
- ☐ HTML format
- ☐ XML format as elements
- ☐ XML format as attributes
- ☐ MS Word Document format

#### **Output Type Selection Control**

## **Task 1.5 - Develop QA/QC System**

The Quality Assurance/Quality Control (QA/QC) standards and processes established for this application provide for multiple layers of quality control. It is important to remember that the standards and processes mentioned in this document do not examine the quality of the data submitted, but rather ensure that the data entering the database is the same data provided by the submitting authority.

As discussed in the first Semi-Annual Technical Report, the Automatic Population Module (APM) requires that data be submitted in a prescribed format. This ensures that each value or flag is positioned in the correct location within the data file and that each item is successfully added to the database. Should any problems with the data be discovered at a later date, then that entire data transaction can be removed, replaced or updated by submitting a replacement data transfer file or removal request to the Application Administrator. This transaction layer is being built to provide reverse-posting capabilities to the application, in addition to the standard transaction processing capabilities provided by Microsoft SQL Server 2000. This layer utilizes a series of timestamp values and log files to associate groups of data with file uploading operations and username activities.

## **B. Phase Two Tasks**

### **Task 2.1 - Develop Stakeholder Website**

As proposed by *ATS*, each stakeholder will have access to the entire data analysis package while the general public will have access to selected features through the public website described in Task 2.2. The stakeholder website will provide the ability to view and develop graphical representation of the digital data online for reports and for data analysis. The data analysis package will be an interactive tool that will be embedded in the data warehouse and repository. The querying of the data permits user-defined access and review of the data. Built-in online analytical tools for advanced data analysis will provide the following options:

- Dynamic/interactive charting capabilities – online graphing of the data in user-defined formats
- Trend analysis – time series of pollutant data – by species, monitor and region
- Back trajectory analysis
- Online point source modeling capabilities
- Multi dimensional plotting capabilities (three dimensions in space (x, y, z), and time)
- Statistical analysis of pollutant profiles and distributions
- Meteorological evaluations (influence on air pollutant concentrations)

The Stakeholder Website is being developed using Microsoft Visual Studio .NET, in conjunction with Microsoft Internet Information Services (IIS), Microsoft SQL Server 2000 and the .NET framework (a packaged addition for MS Windows 2000 or XP). The project team decided to utilize the .NET framework early in the planning stages of this project because of the extensive tool sets available for this platform and the tight integration of XML Web Services into the product. XML Web Services allow remote

users to retrieve datasets locally, combine multiple data sources into a single dataset and exchange data with other datasets that may, or may not, be directly related to the PM<sub>2.5</sub> data. This is crucial to be able to efficiently serve large quantities of data to researchers desiring to query for non-obvious associations between particulate matter data and other data sources, such as health-related data.

As part of the development process, the project team will engage stakeholders to provide critical feedback so the development team can create meaningful and useful analysis tools. The persons and organizations performing air quality data collection and analysis under NETL-sponsored projects are among the most critical stakeholders in this effort; therefore, a meeting to solicit the feedback of these “internal stakeholders” will be held at the NETL Pittsburgh facility in March 2004. Following this meeting, the plan for obtaining feedback from external stakeholders will include a series of interactive web-casts and demonstrations by key personnel from the ATS project team. The web-casts will involve online demonstrations of the website and existing analysis tools.

The initial web-cast will be conducted with a select group of external stakeholders who have previously been made aware of the intent and focus of our database and analytical tool development effort. These stakeholders can provide critical feedback on the types of analytical tools and interfaces which would be most valuable to the research and government/industry communities. This group of stakeholders will be used as a focus group to set priorities and metrics for the development process. A potential list of candidates for the initial web-cast has not been finalized but we have identified the following:

- Randy Hock - Ohio EPA
- S.T. Rao – U. S. EPA ORD
- Donna Kenski - LADCO
- Eric Edgerton, Atmospheric Research and Analysis, Inc.
- Rich Poirot - Vermont Dept. of Environmental Conservation
- Paul Solomon - U.S. EPA, ORD
- Rudolf Husar - Washington University in St. Louis
- Serpil Kayin - MARAMA
- Gary Kleiman - NESCAUM
- Les Hook - NARSTO Quality Systems Science Center

Following this initial web-cast, a set of 2 – 3 follow-up web-casts will be performed as a means of soliciting additional input from the scientific and technical community. It is anticipated that the key stakeholders listed above will identify additional stakeholders who will participate in the follow-up web-casts and provide valuable guidance toward the development of appropriate analytical tools. The follow-up web-casts will also serve to ensure that the project team is implementing the stakeholders’ recommendations satisfactorily.

The second means of securing stakeholder involvement will be to make oral presentations and/or hands-on demonstrations at selected national conferences and meetings on air quality research. The project team is currently considering the possibility of such presentations/demonstrations at the Air & Waste Management Association (AWMA) 97<sup>th</sup> Annual Conference and Exhibition in Indianapolis, IN in June 2004, or at AWMA’s

Visibility Specialty Conference in Asheville, NC in October 2004. It may be possible to set up demonstration rooms equipped with computers for hands on demonstration of the web site at these venues. These demonstrations will be used for feedback, training and to publicize the site to the environmental community.

## **Task 2.2 - Develop Public Website**

*ATS* proposed to construct a separate website connected to the data archive for public outreach, providing citizens and regulatory authorities with a resource and an educational tool highlighting the extensive monitoring programs undertaken by NETL. Publicly accessible sections of the database application's web space will be available to everyone without log in. This portal will be different from the one for stakeholders, sponsors and developers, and will require registered users to submit a username and password combination before access to the restricted website is granted.

This interactive web-based application will be the backbone of the public outreach system. The web delivery system will be designed as an information/decision support center and an educational tool. The system will provide clear and concise data summaries from the monitoring programs and will include easy-to-understand graphical representation of the data including spatial and temporal mapping of the data accompanied by the online help as described in Task 2.3. To insure that the website will deliver information in a clear and concise manner, the deliverables of this task will be reviewed continuously by environmental and community representatives from the region prior to launching.

To date, no items have been developed for the publicly accessible portions of this application. Development is slated to proceed with Task 2.2 in Phase Two of this project.

## **Task 2.3 - Develop Online Help Feature**

*ATS* proposed to construct an online help feature, in conjunction with the web-based application, and it will be developed to support both the sponsor/stakeholder and the public sections website. The online help and instruction component of the application will be an interactive system that will give depth, understanding and context to the environmental data presented. The online help will assist the user at any level of scientific background (novice to professional) in the interpretation of the data. The online help will provide assistance on the following general topics:


- Definitions that will provide clear explanations of the terminology used in evaluating air pollutants
- Explanation of the Federal and State Regulations pertaining to criteria pollutants
- Background information on atmospheric chemistry, transport and emissions of air pollutants
- Effects of meteorology on air pollution episodes
- Significance of the data as it relates to public health
- Information on community-based efforts that can impact ambient air pollution levels
- Navigation of the website itself


To date, only file download help pages have been developed for the publicly accessible portions of this application.




## Download Help


### Download Instructions

 Denotes link to an external site.

 - This attachment is a Word for Windows.

 - This attachment is an Excel spreadsheet.

 - Windows Media Player

 - Microsoft Powerpoint document; When you click on the link an alert box will appear asking you whether you would like to open the Powerpoint document or Save it on your computer. Select open if you want to read it. To save it your PC click on save.

 - This attachment is a Portable Document Format (PDF) document. To view Portable Document Format (PDF) documents identified by a  (PDF icon) you will need to download the free Acrobat PDF reader software.

Once you have the software installed you can do one of the following:

#### Save the document onto your PC (recommended)

Click on the document title with your RIGHT mouse button and then select the option Save as.

This allows you to save the document to the Desktop of your PC. You can then open the document by double clicking on the new PDF Document icon which will have been displayed on your Desktop.

NOTE: You can specify any other location for the file in the Save as dialogue box.

#### Open the document online

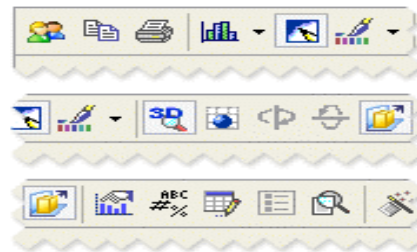
Click on the link with the LEFT mouse button to open the attachment for online viewing. This will open Word, Excel or the the Adobe Acrobat reader so you can read the document.

## Task 2.4 - Provide Graphical Representation of Data

The need for a separate .NET graphing control set became apparent while developing the graphing and analysis tools for this project. The project programmers discussed various charting and graphing packages available and selected ChartFX for .NET. This control set employs both client- and server-side processing functionality. The server-side processes handle the data reading and calculations while the client-side controls offer built in functionality such as save to file, print, zoom, color selection and data editing for *what if* scenarios.

ChartFX creates image files in multiple formats including PNG, JPG and SVG (Scalable Vector Graphics) and stores them, along with a cached dataset, on the server. ChartFX for .NET graphing and charting controls offer potential server-side processing advantages by processing all the analysis and graphic works on a server.

Client-side controls handle user input and send only edits, changes and updates back to the server to reprocess the analysis and graphing request.



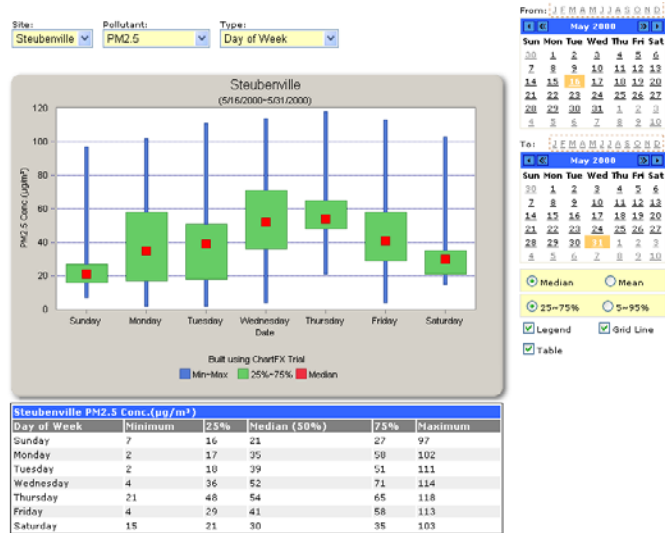
**Example Client-Side Toolbars**

To date, three analysis tools have been developed and published on a web site for stakeholders to view. These are time series plots, pie charts and box-and-whisker plots. Currently site comparison, back trajectory and cluster analysis tools are under development. The project team is actively seeking input from stakeholders regarding desired analysis tools. As new tools are requested, an evaluation and development process will occur to incorporate these tools into the overall toolset.



## Task 2.5 - Performance Testing

*ATS*, in coordination with all members of the internal and external stakeholder groups, will conduct a detailed testing program for the resulting application to verify the functionality and proper execution of all portions of this application. This testing program will provide for interactive user feedback, discussion forums and periodic email notifications and announcements. This testing program will help insure that the intended objectives of this project are met or exceeded. This effort may require revisiting and reworking the original designs, and consequently, will be an ongoing exercise in Phase Two of this project. A section of the website will be dedicated for posting of feedback from beta-testers of this application.



Example Box-Whisker Plot with Data Table

## III. RESULTS AND DISCUSSION

The first phase includes data inventory, benchmarking and database population tasks, as well as the development of data management architecture, a web-based retrieval system and an internal QA/QC system. A progress summary for Phase One is shown in the following table:

Task #	Description	Planned completed %	Actual completed %
1.1	Data Inventory/Benchmarking for Database Applications	100%	95%
1.2	Develop Data Management System Architecture	100%	95%
1.3	Population of Database	100%	65%
1.4	Develop Web-based Retrieval System	100%	90%
1.5	Develop Internal QA/QC System	N/A	N/A

N/A – on-going activity.

The second project development phase will be completed in the second year of the project, which started in August 2003. The tasks involved in this phase include the development of a stakeholder-specific website, a publicly accessible website and an online help feature. This phase also includes the development of special analysis tools to provide a graphical representation of the data and, of course, a series of performance tests designed to provide the best possible data management solution. A progress summary for Phase Two is shown in the following table:

Task #	Description	Planned completed %	Actual completed %
2.1	Develop Stakeholder Website	100%	15%
2.2	Develop Public Website	100%	15%
2.3	Develop Online Help Feature	50%	5%
2.4	Provide Graphical Representation of Data	25%	5%
2.5	Performance Test	25%	0%

## IV. CONCLUSION

The development efforts have so far proceeded as expected. Some target milestones have not been met primarily due to delays in acquiring input data from third party sources. This is especially so with the data inventory task, where data reformatting issues have also been encountered. Delays have also been an inevitable consequence of the philosophy of the DOE COR and the project team to proceed very carefully and deliberately with the development of the stakeholder website. It is believed that the ultimate success of this project will require a high degree of stakeholder confidence and subsequent participation in the website development process, and that such participation will be greatly enhanced if stakeholders are presented with a relatively “polished” product at the outset. Therefore, extra programming effort has been dedicated toward developing and refining a limited set of fully-functional graphic and analytical routines (e.g., time series analysis and box plots) before fully pursuing a potentially expensive program of stakeholder engagement.

The hurdles encountered, however, have not been insurmountable, as other parts of the project have proceeded as proposed. Since the level of effort associated with the slowed tasks is still the same, the costs to complete this task will not be impacted either negatively or positively. Therefore, a request for a twelve-month, no-cost, time extension will be submitted to DOE/NETL in March 2004.

## V. REFERENCES

1. **O'Brien, G., Hughes, V., Modeling Support, California Air Resources Board**  
[Designing a Data Management System for the Central California Air Quality Studies](http://www.arb.ca.gov/airways/Documents/DOCS/AWMACCAQS_Mar2301.pdf)  
[http://www.arb.ca.gov/airways/Documents/DOCS/AWMACCAQS\\_Mar2301.pdf](http://www.arb.ca.gov/airways/Documents/DOCS/AWMACCAQS_Mar2301.pdf)
2. **Professors Spyros Pandis, Cliff Davidson, and Allen Robinson, Carnegie Mellon University**  
[EPA Supersite - Carnegie Mellon University](http://homer.cheme.cmu.edu)  
<http://homer.cheme.cmu.edu>
3. **Robinson P. Khosah, Ph.D., Advanced Technology Systems, Inc.**  
[Semi-Annual Technical Progress Report](http://www.netl.doe.gov/coalpower/environment/air_q/docs/40456r06.pdf)  
[http://www.netl.doe.gov/coalpower/environment/air\\_q/docs/40456r06.pdf](http://www.netl.doe.gov/coalpower/environment/air_q/docs/40456r06.pdf)
4. **Karen Magliano, California Air Resources Board; Philip Roth, Envair; Charles Blanchard, Envair; Steven Reynolds, Envair; Steve Ziman, Chevron; Rob DeMandel, Bay Area Air Quality Management District**  
[California Regional Pm<sub>10</sub>/Pm<sub>2.5</sub> Air Quality Study: Objectives and Associated Data Analysis and Modeling Approaches](http://www.arb.ca.gov/airways/Documents/reports/objectiv.doc)  
<http://www.arb.ca.gov/airways/Documents/reports/objectiv.doc>
5. **Robinson P. Khosah, Ph.D., Charles G. Crawford, Advanced Technology Systems, Inc.**  
[Semi-Annual Technical Progress Report](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R01.pdf)  
[http://www.netl.doe.gov/coalpower/environment/air\\_q/docs/41476R01.pdf](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R01.pdf)
6. **Robinson P. Khosah, Ph.D., Charles G. Crawford, Advanced Technology Systems, Inc.**  
[Semi-Annual Technical Progress Report](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R02.pdf)  
[http://www.netl.doe.gov/coalpower/environment/air\\_q/docs/41476R02.pdf](http://www.netl.doe.gov/coalpower/environment/air_q/docs/41476R02.pdf)

## **VI. BIBLIOGRAPHY**

Not applicable.

## VII. LIST OF TERMS, ACRONYMS AND ABBREVIATIONS

Term	Definition
Admin Level	Security level indicating the degree of access a specific user possesses to administration utilities and data.
ACHD	Allegheny County Health Department
APM	Automated Population Module
Application Account	An application account (Windows 2000). This account is created and edited using Windows 2000.
Application Administrator	An individual responsible for managing application performance, user access and newsletter/announcement notification services.
Browse Level	Security level indicating the areas of the application and datasets that a specific user can 'browse' through.
Cached Data	Data retained at the server level to serve frequently polled data. These datasets are cached, or stored, at the server and reduce query loads on the database server, thus increasing overall efficiency and application response time.
CARB	California Air Resources Board
CSV	Comma Separated Value
Data Administrator	An individual responsible for managing the database housing the PM data, as well as managing all assigned data submission accounts.
Data Submitter	Individual user who has permission to submit data for inclusion in the PM database.
DOE-NETL	US Department of Energy's National Energy Technology Laboratory
Foreign Key	A non-negative whole number used to reference a data row in a related table.
FTP	File Transfer Protocol.
GIS	Geographic Information Systems
GMT Offset	Number of hours that, when added to the local time values, provides GMT Time values; e.g. 11:00AM local time, with a GMT offset value of -5 means that the GMT time value for this local time value would be 6:00AM GMT.

HTTP	Hypertext Transfer Protocol
Media	Filter used to collect speciation samples.
Metaflag	Localized flagging system specific to a particular submitting authority.
Method	Descriptive text that describes how data was collected.
NARSTO	An acronym for "North American Research Strategy for Tropospheric Ozone." A tri-national, public-private partnership for dealing with multiple features of tropospheric pollution, including ozone and suspended particulate matter.
NARSTO Metaflag	Standardized flagging system (NARSTO). Each metaflag is mapped to a NARSTO metaflag to provide meaningful results when querying across datasets originating from multiple submitting authorities.
NOAA	National Oceanic and Atmospheric Administration
PA-DEP	Pennsylvania Department of Environmental Protection
Parameter	A concatenated descriptive definition of what the observation value represents. Components of a valid parameter include: parameter property, parameter identifier, collection principle, and parameter source.
Parameter Identifier	Descriptive text that identifies a chemical property of a parameter.
Parameter Property	Descriptive text that identifies a physical property of a parameter.
Primary Key	Unique non-negative whole number used to reference each row in a database table. This is used to identify relationships between related items in related tables.
Parameter Source	Originating organization for parameter codes and descriptions.
QA/QC	Quality Assurance / Quality Control
QC Status	Quality control status code.
Read Level	Security level indicating the areas of the application and datasets to which a specific user has read access.
Sample Duration	Text describing the sample duration that is used to collect a specific sample. This usually applies only to filter data; sample duration of H12 indicates that the sample in question was taken over a 12-hour period.
Sample Frequency	Text describing the sample frequency, or interval, between regular readings; e.g. M15 indicates that a sample is taken every 15 minutes.

SQL	Structured Query Language
Subscriber	Individual user who has elected to receive email notification from pmdata.org.
System Account	A Windows 2000 Server account used to administer the network and/or application servers.
Systems Administrator	An individual responsible for managing the hardware and operating system(s) of the hosting computers and networks. This person ensures that the application and database is available to users and works to correct any connectivity issues that may occur.
User Account	Application account established for each user that contains each user's contact data and security profile.
US EPA	US Environmental Protection Agency
VCard	Virtual address card. This is similar to a rolodex entry, containing an address, city, state and zip code. A VCard can link to multiple entities sharing the same physical address. Entries also contain a location's county and country.
Write Level	Security level indicating the areas of the application and datasets to which a specific user may enter new records or modify existing records.
WV-DEP	West Virginia Department of Environmental Protection
XML	Extended Markup Language